Available online at www.sciencedirect.com

SCIENCE DIRECT®

Virology 322 (2004) 239–252

VIROLOGY

www.elsevier.com/locate/yviro

Genomic organization and molecular analysis of the inducible prophage EJ-1, a mosaic myovirus from an atypical pneumococcus

Patricia Romero, Rubens López, and Ernesto García*

Departamento de Microbiología Molecular, Centro de Investigaciones Biológicas, CSIC, 28040 Madrid, Spain

Received 15 December 2003; returned to author for revision 09 January 2004; accepted 31 January 2004

Abstract

We report the complete genomic sequence of EJ-1, an inducible prophage isolated from an atypical *Streptococcus pneumoniae* strain that belongs to the *Myoviridae* morphology family. The phage and bacterial recombinational sites (attachment sites) have been also determined. The genome of the EJ-1 prophage (42935 bp) is organized in 73 open reading frames (ORFs) and in at least five major clusters. Bioinformatic and N-terminal amino acid sequence analyses enabled the assignment of possible functions to 52 ORFs. The predicted proteins coded for the EJ-1 genome revealed similarities in the lysogeny, DNA replication, regulation, packaging, and head morphogenesis protein clusters with those from several siphoviruses infecting lactic acid bacteria. However, the proteins encoded by genes *orf53* to *orf64*, corresponding to putative tail proteins of the virion, were very similar to those of the defective *Bacillus subtilis* myovirus PBSX with the notable exception of the gene product of *orf56* (the tape measure tail protein) that was similar to proteins from phages infecting Gram-negative bacteria. The first description of the genome of a myovirus infecting a low G + C content Gram-positive bacterium, a member of a group embracing important human pathogens and industrial relevant species, will contribute to expand our current knowledge on phage biology and evolution.

Keywords: Bacteriophage evolution; Comparative genomics; Genetic mosaic; *Streptococcus pneumoniae*; *Myoviridae*

Introduction

During the last few years, the easy and low cost of nucleotide (nt) sequencing has allowed the determination of more than 100 complete phage genome sequences. Among those infecting low G + C content Gram-positive (LGP) bacteria, much data are currently available from *Siphoviridae* phages infecting dairy bacteria (Brüssow, 2001; Brüssow and Desiere, 2001; Brüssow and Hendrix, 2002; Desiere et al., 2002). The recent sequencing projects involving streptococcal and lactococcal genomes have also provided additional insights on the evolution of phage genomes (Desiere et al., 2001a, 2001b). Nevertheless, almost all the entirely sequenced phage genomes belong to siphoviruses whereas only few complete podoviral genomes infecting LGP bacteria have been reported (Martín et al., 1996; Meijer et al., 2001; Nelson et al., 2003; Vybiral et al., 2003).

The current taxonomic system approved by the International Committee on the Taxonomy of Viruses (ICTV) (http://www.ncbi.nlm.nih.gov/ICTV) presently classifies bacteriophages, following a hierarchical arrangement, into one order, 13 families, and 30 genera (Ackermann, 2003). This classification is based on the morphology of virions and on features of their nucleic acid as being DNA or RNA, double- or single-stranded, etc. Besides, the wealth of phage genome sequencing projects (there are now about 150 genome sequences of tailed phages in the databases, and this number is increasing rapidly) has revealed that phages represent a good example of genetic mosaicism apparently arising from nonhomologous recombination between ancestral sequences (Hendrix, 2003) following a web-like, rather than a tree-like, phylogeny (Brüssow and Hendrix, 2002). This finding has revealed that the official taxonomy of phage species no longer reflects our knowledge on phage biology and evolution, and several authors have recently proposed different models based on genome sequences

* Corresponding author. Departamento de Microbiología Molecular, Centro de Investigaciones Biológicas, CSIC, Ramiro de Maeztu 9, 28040 Madrid, Spain. Fax: +34-91-536-0432.

E-mail address: e.garcia@cib.csic.es (E. García).

0042-6822/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2004.01.029

(Lawrence et al., 2002; Proux et al., 2002; Rohwer and Edwards, 2002). Unfortunately, it is still difficult to reach definite conclusions since (1) we possess only a miniscule sample of complete phage genomes; (2) most sequences are from only three bacterial systems, that is, coliphages, dairy phages, and mycobacteriophages; and (3) most of the studied phages whose genome is known correspond to those having long, noncontractile tails (*Siphoviridae*).

EJ-1 is a temperate bacteriophage belonging to the *Myoviridae* family (phages with long, contractile tails) (Fig. 1). This phage was purified from mitomycin C-induced cultures of an atypical *Streptococcus pneumoniae* strain named 101 that was isolated in 1987 at the Hospital de San Juan de Dios (Barcelona, Spain) from the blood of a patient suffering from pneumonia (Díaz et al., 1992a). Although this strain was resistant to bile and optochin and nontypeable, it was shown to harbor a *lytA* allele ($lytA_{101}$) coding for a bile-sensitive lytic enzyme (Díaz et al., 1992b; Obregón et al., 2002), a gene characteristic of pneumococcal isolates (Kawamura et al., 1999). In addition, EJ-1 has a lytic gene (*ejl*) very similar to $lytA_{101}$ (81% identity), whose product is also a bile-sensitive lytic enzyme, and a holin gene (*ejh*) required to allow the Ejl lytic amidase reaching its peptidoglycan target (Díaz et al., 1996; Haro et al., 2003; Sáiz et al., 2002). Recent works strongly suggest that phages might provide fundamental clues to understand global diversity at the species and genome level and to study the recruitment of interesting genes in their genomes. This is particularly important in pathogenic microorganisms like *S. pneumoniae* where a high presence of temperate pages has been repeatedly suggested (Bernheimer, 1977; Obregón et al., 2003a,b; Ramirez et al., 1999; Severina et al., 1999). Furthermore, myoviruses represent a group of phages for which only a few sequence genomes have been dissected and warrant further investigation. To the best of our knowledge, this report represents the first account on the complete nt sequence of a genome from a myovirus infecting LGP bacteria.

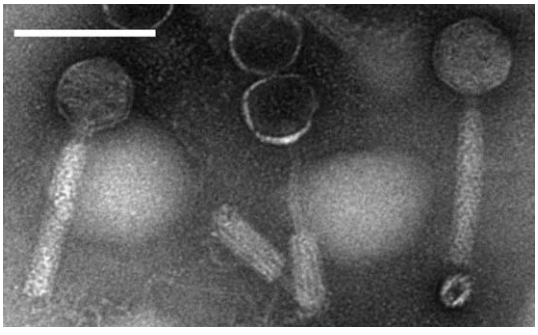

Fig. 1. Electron micrograph of a purified preparation of bacteriophage EJ-1 negatively stained with 1% uranyl acetate. Scale bar represents 100 nm.

Results and discussion

Phylogenetic position of the lysogenic, atypical pneumococcal strain 101

From a clinical perspective, it is important to distinguish accurately between pneumococcal isolates and other α-hemolytic streptococci because this diagnosis determines the choice of the corresponding antibiotic treatment. In addition, the potential therapeutic value of several pneumococcal phage lytic enzymes has been recently documented (Jado et al., 2003; Loeffler et al., 2001), but it has been observed that these enzymes showed species specificity. Then, it appeared interesting to further identify the EJ-1 host strain. A partial characterization of the atypical pneumococcal strain 101, which included the sequencing of alleles $lytA_{101}$, *gki*, and *xpt*, has been recently reported (Díaz et al., 1992b; Obregón et al., 2002). To get a more complete determination of the phylogenetic position of this clinical isolate, we have now found that a partial 16S rDNA sequence from strain 101 showed a similarity higher than 99% with strains of *S. pneumoniae*, including strains TIGR4 (Tettelin et al., 2001) and R6 (Hoskins et al., 2001), *Streptococcus mitis*, and *Streptococcus oralis* (data not shown). The pairwise evolutionary distances (PEDs) calculated by sequence comparison between the seven housekeeping genes from the 101 strain and those from pneumococcal strains (Table 1) indicated that strain 101, from the evolutionary viewpoint, is as distant (or close) from pneumococci as it is from *S. mitis*. A similar conclusion was reached when the $galU_{101}$ allele was compared to those included in the data banks (data not shown). Additional tests (not shown) also gave conflicting results, that is, the sequence of the $sodA_{101}$ allele and the absence of the pneumolysin gene (*pnl*) indicated that the 101 strain is a *S. mitis* isolate but it lacks the *ant* gene, characteristic of *S. mitis* and *S. oralis* isolates (Balsalobre et al., 2003). These analyses indicate that EJ-1 has been isolated from a strain that is located taxonomically at the edge between clinically outstanding LGP bacteria, a peculiarity that makes the study of the genome of EJ-1 an appropriate reference for furthering our understanding on genetic interchanges among phages, and between phage and bacteria. According to the current rules (Stackebrandt et al., 2002), DNA–DNA hybridization experiments appear to be mandatory to definitely answer to this question.

Genome analysis of prophage EJ-1

The criteria used to identify putative open reading frames (ORFs) were (1) that they could code for a polypeptide of more than 33 amino acid (aa) residues, and (2) that they showed a putative ribosome binding site (RBS) (Bacot and Reeves, 1991) and, at a convenient distance, began with an ATG, GTG (three ORFs), or TTG codon (six ORFs). Using these criteria, our analyses revealed the

Table 1
Pairwise comparison of the seven housekeeping alleles of strain 101 with those included in the pneumococcal MLST database

Strain	Pairwise evolutionary distances[a]																	
	aroE			*ddl*		*gdh*		*gki*			*recP*			*spi*		*xpt*		
	Spn[b] ($n = 56$)	Smi[c]	Smi[T,d]	Spn ($n = 124$)	Smi	Spn ($n = 78$)	Smi	Spn ($n = 88$)	Smi	Smi[T]	Spn ($n = 59$)	Smi		Spn ($n = 83$)	Smi	Spn ($n = 126$)	Smi	Smi[T]
Spn[b]	≤ 2.0	0.2–1.5	4.1–5.4	≤ 11.8	4.9–10.0	≤ 6.6	0.4–4.5	≤ 7.4	3.6–6.7	3.2–7.6	≤ 5.5	0.2–4.8		≤ 9.0	0.2–7.3	≤ 9.4	0.4–7.3	4.9–9.9
Smi[c]	0.2–1.5	–	4.3	4.9–10.0	–	0.4–4.5	–	3.6–6.7	–	4.7	0.2–4.8	–		0.2–7.3	–	0.4–7.3	–	7.3
Smi[T,d]	4.1–5.4	4.3	–	ND[e]	ND	ND	ND	3.2–7.6	4.7	–	ND	ND		ND	ND	4.9–9.9	7.3	–
101	5.6–6.7	5.4	5.4	0.0–9.4	7.4	3.3–4.5	3.8	2.1–6.0	3.6	5.6	4.3–5.8	5.1		4.3–7.5	6.1	4.9–10.3	7.3	5.1

[a] PEDs were calculated using the program DISTANCES and the sequences of all pneumococcal alleles included in the MLST database (http://www.spneumoniae.mlst.net/) (3 October 2003, last date accessed). Numbers represent the estimated number of substitutions per 100 bases using the Jukes–Cantor distance correction (Jukes and Cantor, 1969). *n*, number of alleles.
[b] The values correspond to the pneumococcal strains (Spn) included in the MLST database.
[c] The corresponding alleles of this strain (PN92/120) are also included in the MLST database.
[d] Type strain (ATCC 49456 = NCTC 12261). The accession numbers are *aroE*, AY283586; *gki*, AY292542; *xpt*, AJ422286.
[e] ND, not done because no sequence was available.

presence of 73 ORFs in a unit genome size of 42 935 bp in the prophage state (Fig. 2). The average G + C content of EJ-1 DNA was calculated as 39.6%, which is practically identical to the 39.7% reported for the *S. pneumoniae* genome (Tettelin et al., 2001) (Fig. 2). No evidence could be found for the presence of cohesive, protruding ends (*cos*) in the EJ-1 genome by treatment with DNA ligase (unpublished observations). Moreover, restriction endonuclease digestion of EJ-1 DNA also failed to reveal a submolar fragment possibly containing a *pac* site, although a similar behavior has been reported for other circularly permuted phage genomes such as A118 of *Listeria monocytogenes* (Loessner et al., 2000), and the coliphage 933W (Plunkett et al., 1999). Besides, time-limited treatment of EJ-1 DNA with the exonuclease BAL-31, followed by complete digestion with restriction enzymes, revealed that all fragments were simultaneously degraded, in contrast to the specific truncation of fragments observed in the control, λ DNA (results not shown). These results taken together suggested that there are no invariable ends in the mature EJ-1 DNA molecules, that is, the packaged DNA is circularly permuted.

The nt sequence of the right end of the prophage (an *Eco*RI–*Bcl*I DNA fragment of 1903 bp) that includes the holin (*ejh*) and the lytic amidase gene (*ejl*) as well as the *attR* region and its downstream sequence had been previously reported (Díaz et al., 1992a). Bioinformatic analyses indicated that the predicted gene products (gp) of 52 ORFs showed significant sequence similarities with proteins included in the databases (Table 2). The genetic organization of left part of the EJ-1 prophage (about 12-kb long) (from *orf1* to *orf24*) shows the greatest global similarity (at the aa level) to the *Streptococcus pyogenes* defective prophage SF370.3 (Desiere et al., 2001b) (Table 2 and Fig. 2). The lysogeny module shows a gene order characteristic of temperate phages from LGP bacteria (*attL*-integrase-unknown-*cI*-like repressor-*cro*-like repressor-excisionase), although this conserved gene order is apparently disrupted by inserted genes (e.g., *orf11* and *orf12*) that appear to be less related to SF370.3 than to other different phages (Table 2 and Fig. 2). The putative excisionase gene (*orf16*) is located far apart from the lysogeny module being located in the replication cluster. At least four different lysogeny modules have been observed in phage genomes including the λ supergroup of *Siphoviridae* and the P2-like genus of *Myoviridae* (Brüssow and Desiere, 2001). However, EJ-1 appears to represent a peculiar case since the distance between the genes coding for the integrase and the excisionase, which usually ranges from zero to six ORFs both in Gram-positive and Gram-negative temperate bacteriophages, or even the excisionase gene, is missing as in phage Sfi21 (Bruttin et al., 1997); in the EJ-1 prophage, the corresponding genes are *orf1* and *orf16*, respectively. This finding suggests the insertion in the EJ-1 genome of intervening genes from a foreign source (as suggested

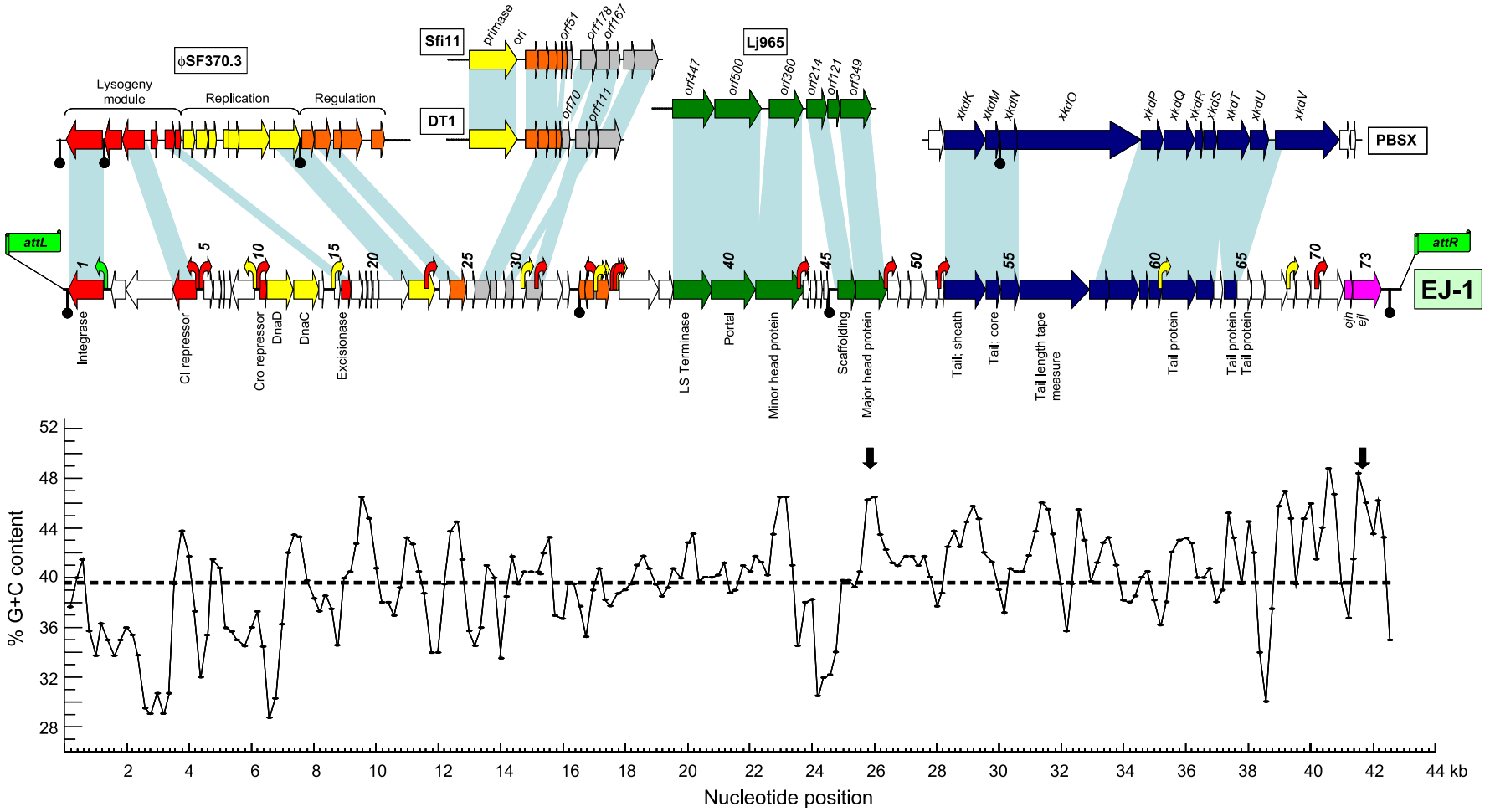

Fig. 2. Genomic organization of the EJ-1 prophage and comparison with other phage genomes infecting LGP bacteria. ORFs (arrows) were colored according to their predicted function: red, lysogeny; yellow, replication; orange, transcriptional regulation; dark green, head morphogenesis; dark blue, tail components; pink, host lysis; gray, putative regulatory genes, open, hypothetical ORFs. , extended and other putative canonical promoters, respectively. , putative transcriptional terminators. Light blue shading connects genes that are linked by significant sequence similarities between phage genomes. At the bottom, G + C content of the EJ-1 genome. A window of 400 nt (200 nt step) was used. The dotted line represents the average % G + C content of the EJ-1 genome. Vertical arrows indicate the position of two G + C-rich (> 50%) islands longer than 200 bp (nt 25770–26026, and 41584–41784).

Table 2
Comparative analysis of the genes from *S. pneumoniae* phage EJ-1 with proteins included in the databases

Gene	Gene product				Related phage and bacterial proteins		
orf	Start	Stop	aa	Putative function (conserved domains)	Protein (origin)[a] (% aa identity)	$\log_{10}E$ value (aa overlap)	Accession no.
1	1248	106	380	integrase	prophages 370.3/370.3-like/315.3/SPsP4; putative integrase (Spy) (74.7)	−116 (380)	Q99YY3
					phage MM1; integrase (Spn) (33.1)[b]	−35 (378)	Q94M77
2	1943	1497	148	–	–	–	–
3	3500	1959	513	HP[c] (COG0711)	prophage ϕSDA; HP (Spy) (53.9)	−67 (503)	Q8NZP2
4	4332	3529	267	CI repressor (COG2932; PF01381)	prophages 370.3/315.3/370.2/315.2/SPsP4; putative repressor (Spy) (44.9)	−29 (237)	Q99YY5
5	4508	4783	91	–	–	–	–
6	4856	4999	47	–	–	–	–
7	5061	5165	34	–	–	–	–
8	5170	5364	64	–	–	–	–
9	6099	5326	257	HP	phage SM1, Gp14 (Smi) (50.8)	−31 (182)	Q7Y4L0
10	6308[d]	6535	75	Cro repressor (COG1396; PF01381)	prophage 315.3; HP (Spy) (64.3)	−12 (70)	Q8K761
11	6532	7380	282	DnaD (COG3935; PF04271)	phage 7201; Orf4 (Sth) (31.4)	−12 (241)	Q9MCM7
					plasmid pSX267; replication initiator protein RepA; (Sxy) (30.6)[b]	−9 (219)	Q52583
12	7392	8204	270	replication protein (COG1484; PF01695)	prophage ϕspeA; putative DNA replication protein (Spy) (35.4)	−26 (251)	Q8P2I1
13	8201[e]	8371	56	–	–	–	–
14	8385	8528	47	–	–	–	–
15	8725[d]	8901	58	transcriptional regulator (COG3655; PF01381)	prophage ϕspeLM; HP (Spy) (55.2)	−8 (58)	Q8P0L9
16	8980	9291	103	excisionase	prophages 370.3/370.3-like/315.5/SPsP4; putative excisionase (Spy) (74.8)	−30 (103)	Q99YY7
17	9303	9650	115	–	–	–	–
18	9667	9810	47	HP	prophage ϕspeC; HP (Spy) (46.7)	−4 (45)	Q8P1S5
19	9797	9919	40	–	–	–	–
20	9916	10,095	59	–	–	–	–
21	10,092	11,063	323	HP (COG3723; PF03837)	prophage Lp1; protein 18 (Lpl) (45.9)	−40 (303)	Q88YV1
22	11,067	11,903	278	HP	prophage 370.3; HP (Spy) (37.2) prophages 315.3/SPsP4,	−21 (266)	Q99YZ3
23	12,078	12,410	110	HP	HP (Spy) (61.5)	−26 (109)	Q8K6U2
24	12,407	12,922	171	HP	prophage 370.3; conserved HP (Spy) (78.0)	−51 (168)	Q99YZ5
25	12,934	13,215	93	–	–	–	–
26	13,223[e]	13,708	161	HP	phage Sfi11; HP (Sth) (50.0)	−14 (132)	Q9MCI0
27	13,709	13,918	69	HP	phage DT1; HP (Sth) (33.8)	−4 (69)	Q9XJD4
28	13,918	14,193	91	–	–	–	–
29	14,190[e]	14,450	86	HP (PF07116)	phage SM1, Gp27 (Smi) (65.1)	−21 (86)	Q7Y4J7
30	14,447	14,848	133	–	–	–	–
31	14,857	15,354	165	DNA-binding protein	conserved HP (Sag) (35.0)	−12 (162)	AAN00730
32	15,477	16,037	186	transcriptional regulator	HP (Lin) (31.3)	−10 (174)	Q926A7
33	16,039	16,263	74	HP	phage MM1; HP (Spn) (93.1)	−26 (74)	Q8SBI6

(*continued on next page*)

Table 2 (*continued*)

Gene	Gene product				Related phage and bacterial proteins		
orf	Start	Stop	aa	Putative function (conserved domains)	Protein (origin)[a] (% aa identity)	$\log_{10}E$ value (aa overlap)	Accession no.
34	16,636[e]	16,415	73	transcriptional regulator (COG2337; PF02452)	cell growth regulatory protein (Lpl) (70.8)	−11 (48)	Q88TP7
35	16,907	16,620	95	transcriptional regulator	transcriptional regulator, AbrB family (Efa) (46.0)	−4 (51)	AAO82928
36	16,965[d]	17,411	148	transcriptional regulator	phages NIH1.1/ϕseLM/315.4/SPsP3; HP (Spy) (37.6)	−13 (140)	Q938L8
37	17,847[d]	19,112	421	transferase (COG0863; PF01555)	HP (Sag) (47.3)	−74 (420)	Q8E5B2
38	19,142	19,603	153	HP	phage 370.1; HP (Spy) (52.7)	−18 (127)	Q9A0P5
39	19,591[e]	20,856	421	LS Terminase (COG1783; PF03237)	prophage terminase large subunit (Cte) (59.9)	−103 (413)	AAO35668
40	20,861[e]	22,276	471	portal (PF05133)	phage SPP1; portal protein (Bsu) (31.5)[b]	−31 (425)	P54309
41	22,257	23,801	514	minor head protein (COG2369; PF04233)	phage SM1, Gp35 (Smi) (41.8)	−30 (236)[f]	Q7Y4I9
42	23,764	23,973	69	HP	HP Nma0318 (Nme) (86.2)	−20 (58)	Q9JWL2
43	24,027	24,200	57	HP	phage MM1; HP (Spn) (36.8)	−3 (57)	Q94M48
44	24,226	24,378	50	–	–	–	–
45	24,467[e]	24,607	46	–	–	–	–
46	24,961	25,542	193	scaffolding (PF06810)	putative scaffolding protein (Lmo) (31.6)	−11 (193)	Q8Y4Y2
					phage SPP1; scaffolding protein (Bsu) (29.2)[b]	−5 (111)	Q38580
47	25,542	26,534	330	major head	HP Lmo2296 (Lmo) (58.7)	−78 (328)	Q8Y4Y3
					phage SPP1; coat protein (Bsu) (39.9)[b]	−41 (323)	Q38582
48	26,550	26,945	131	–	–	–	–
49	26,952	27,299	115	HP	HP (Lga) (29.0)	−7 (108)	ZP_00046446
50	27,296	27,787	163	HP	–	–	–
51	27,777	28,193	138	HP	conserved protein (Cte) (29.0)	−2 (119)	AAO36622
52	28,186	28,356	56	–	–	–	–
53	28,356	29,669	437	sheath; tail	HP Lin1278 (Lin) (38.8)	−49 (437)	Q92CB1
					phage-like element PBSX; XkdK (Bsu) (39.9)	−18 (419)	P54331
54	29,683	30,147	154	core; tail	phage-like element PBSX; XkdM (Bsu) (44.5)	−17 (146)	P54332
55	30,163	30,759	198	HP	phage-like element PBSX; XkdN (Bsu) (30.8)	−12 (179)	P54333
56	30,763	33,009	748	tail length tapemeasure (COG5281)	phage-related protein Rsc0873 or Rsc04872 (Rso) (27.3)	−12 (381)[g]	Q8Y120
					phage HK97; tail length tape measure protein (Eco) (31.4)	−9 (306)[g]	Q9MCS3
57	33,023	33,715	230	HP	HP Lin1283 (Lin) (32.5)	−10 (153)[g]	Q92CA7
					phage-like element PBSX; XkdP (Bsu) (31.8)	−9 (148)[g]	P54335
58	33,712	34,680	322	HP	HP Lin1284 (Lin) (26.8)	−11 (285)	Q92CA6
					phage-like element PBSX; XkdQ (Bsu) (27.6)	−9 (258)	P54336
59	34,680	34,982	100	HP	HP Lgas0622 (Lga) (29.0)	−6 (98)	ZP_00046455
					phage-like element PBSX; XkdR (Bsu) (27.0)	−1 (89)	P54337
60	34,985	35,377	130	HP	phage-like element PBSX; XkdS (Bsu) (38.3)	−8 (122)	P54338
61	35,367	36,494	375	tail protein (COG3299)	HP Lin1287 (Lin) (31.7)	−29 (361)	Q92CA3

Table 2 (*continued*)

Gene	Gene product				Related phage and bacterial proteins		
orf	Start	Stop	aa	Putative function (conserved domains)	Protein (origin)[a] (% aa identity)	$\log_{10}E$ value (aa overlap)	Accession no.
					phage-like element PBSX; XkdT (Bsu) (33.6)	−17 (366)	P54339
62	36,494	37,087	197	HP	prophage Lj928; HP (Ljo) (32.2)	−8 (149)	Q9AZ88
					phage-like element PBSX; XkdU (Bsu) (20.2)	−3 (183)	P54340
63	37,080	37,355	91	–	–	–	–
64	37,348	37,797	149	tail fiber	phage-like element PBSX; XkdV (Bsu) (50.0)	−4 (60)[g]	P54341
65	37,802	38,290	162	–	–	–	–
66	38,299	38,724	141	–	–	–	–
67	38,731	39,450	239	HP	phage SM1; HP (Smi) (39.4)	−15 (198)	Q9AF63
68	39,462	39,740	92	HP	prophages 315.6/SPsP1; HP (Spy) (29.9)	−4 (78)	Q8K6B7
69	39,755	40,231	158	HP	prophages 315.5/SPsP2 (Spy) (39.8)	−10 (108)	Q8K6K1
70	40,238	40,558	106	–	–	–	–
71	40,564	41,280	238	HP	see reference Díaz et al., 1992a	–	Q38655
72	41,296	41,553	85	holin	see reference Díaz et al., 1992a [b]	–	Q38656
73	41,555	42,505	316	amidase	see reference Díaz et al., 1992a [b]	–	Q38658

[a] Abbreviations: Bsu, *Bacillus subtilis*; Cte, *Clostridium tetani*; Eco, *Escherichia coli*; Efa, *Enterococcus faecalis*; Lga, *Lactobacillus gasseri*; Ljo, *Lactobacillus johnsonii*; Lpl, *Lactobacillus plantarum*; Lin, *Listeria innocua*; Lmo, *Listeria monocytogenes*; Nme, *Neisseria meningitidis*; Rso, *Ralstonia solanacearum*; Sxy, *Staphylococcus xylosus*; Sag; *Streptococcus agalactiae*; Smi, *Streptococcus mitis*; Spn, *Streptococcus pneumoniae*; Spy, *Streptococcus pyogenes*; Sth, *Streptococcus thermophilus*.
[b] Experimental evidence.
[c] HP, hypothetical protein.
[d] Orf starting with a GTG codon.
[e] Orf starting with a TTG codon.
[f] Similarity restricted to the C-terminal moiety.
[g] Similarity restricted to the N-terminal part of the protein included in the database.

by their low G + C content) (Fig. 2), and genome rearrangements).

The gene product of *orf11* (gp11) showed significant similarity to the product of *orf4* (replication initiator protein, DnaD) of the *Streptococcus thermophilus cos* site phage ϕ7201 (Table 2) (Stanley et al., 2000). It is well known that the origin of replication of phages from lactic acid bacteria typically consist of, or contain, direct DNA repeats. On the basis of this criterion, a putative origin of replication may be identified within *orf11*. A 38-bp perfect direct repeat was found (not shown). Moreover, this repeat is 79% identical to the 34-bp repeat located at the ϕ7201 origin. Other shorter repeats were also detected in both phages at similar positions. As a whole, there is 66% nt identity between EJ-1 and ϕ7201 genomes in the 250 bp region around the repeats. In addition, the region containing the repeats has a G + C content of 26%, well below the average of the whole EJ-1 genome. Sequence similarities with the prophage SF370.3 genome disappear downstream of *orf25*. Remarkably, gp26 is 50% identical to the product of *orf178* located at the right end of the *S. thermophilus pac*-site lytic phage Sfi11 genome (Brüssow and Desiere, 2001) (Table 2), whereas gp31 is similar to the product of *orf167*. In the Sfi11 genome, the *orf178* and *orf167* genes are located 3′ from *orf51* that has been recognized as a recombination hot spot (Lucchini et al., 1999). This might also be the case in the EJ-1 genome as deduced for the presence of additional genes. In fact, gp27 and gp29 share significant similarities with the predicted products of two genes (*orf70* and *orf111*, respectively) of phage DT1, a *cos*-site relative of Sf11, that map at a corresponding position in DT1 genome (Fig. 2). Several sequences with potential promoter activity are also present downstream of *orf30* (Table 3). The general organization of the region embracing from *orf39* to *orf48* showed a one-to-one correspondence to similar-size genes in the DNA packaging and head morphogenesis clusters of various *pac*-site, Sfi11-like phages (Desiere et al., 2000) with the exception of an insertion of four small ORFs (*orf42* to *orf45*). Moreover, significant aa similarities between the EJ-1 gene products and those from phage LJ965 are evident (Fig. 2).

The gene cluster putatively encoding the tail components of EJ-1 (from *orf53* to *orf64*) is different from all other previously studied siphoviruses infecting LGP but very similar in gene content and order to that of PBSX (from XkdK to the N-terminal end of XkdV) (Krogh et al., 1996), a defective myovirus of *Bacillus subtilis*, with the noticeable exception of gp56 that may be involved in the determination of the tail length. However, a Western blot analysis carried out using an anti EJ-1 antiserum did not reveal any reaction with the structural proteins of PBSX (not shown). Notably,

Table 3
Predicted canonical and extended promoters in the EJ-1 genome

Position	Sequence (5′ →3′)[a]			Index[b]
1492–1464 (c)[c]	TTGACA	gtatttataaatTgaac	TATAAT	6,1,6
4389–4361 (c)	TTtACA	aatacgaaaataaacGt	TATtAT	5,1,5
4431–4459	TTGACA	tacgatttaattcgTag	TATAAT	6,1,6
6156–6128 (c)	TTctaA	ccacctttcaaaTgTGg	TATAAT	3,3,6
6140–6170	TTGAaA	ggtggttagaattggaaGa	TAaAAT	5,1,5
8663–8692	aTaACA	atatgtagtgttTtTtg	TATAtT	4,2,5
11,584–11,612	TTaACA	atattcttaactTtgGa	TATcAT	5,2,5
14,748–14,778	agGACA	cgatgctagatgtaTcTGa	TtTAAT	4,3,5
15,179–15,226	TTGACt	ctattgattggTcatc	TAaAAT	5,1,5
16,921–16,950	gTGAtA	tacccgataaaaaTaTta	TATcAT	4,2,5
16,969–16,941 (c)	cTcACA	acatcgctccttTaTGa	TATAAT	4,3,6
17,361–17,389	aaaAgg	tagctgattgtcTaTcg	TAaAAT	1,2,5
17,417–17,446	TaGgaA	atggcttaatttacaTGg	TAaAAT	3,2,5
17,470–17,497	TTGACg	tctcctttataTtTat	TATAtT	5,2,5
23,736–23,765	TTGAaA	ccgaaaaaaggagTtgat	TATtAT	5,1,5
26,524–26,554	TTGgCA	aatagattgggtaacagaa	TATgAT	5,0,5
35,310–35,337	TTGAag	tccataccataTcTGg	TaTtAT	4,3,4
39,441–39,471	TaacCA	ataatggaggtgacTtTGg	TAgAAT	3,3,5
40,307–40,434	TTGACA	cactgaaagagatTGc	TAcAAT	6,2,5

[a] Bases identical to consensus are indicated in uppercase lettering.
[b] Bases identical to consensus in −35, −10 extension, and −10 boxes; maximum is 6,3,6.
[c] (c) indicates that the promoter is located in the complementary strand to that included in the data bank.

the predicted tail length tape measure protein of EJ-1 (gp56) showed significant similarities with the N-terminal moieties of proteins from several phages infecting Gram-negative bacteria including the lambdoid or λ-like coliphage HK97 (Table 2) (see below). Interestingly, and in sharp contrast, sequence similarities between gpH proteins encoding the tail length tape measure proteins of phages λ and HK97 only extend over to the C-terminal regions of the proteins (Juhala et al., 2000).

Several genes encoding hypothetical proteins of unknown function are located further downstream. Besides, gp72 and gp73 corresponding, respectively, to the Ejh holin and the Ejl lytic amidase have already been investigated in detail (Díaz et al., 1992a, 1996; Haro et al., 2003; Sáiz et al., 2002).

Determination of the phage EJ-1 integration site

The *attR* region of the EJ-1 prophage is located downstream of *ejl* (*orf73*) (Díaz et al., 1992a). The sequence of *attL*, the hybrid site to the left of the prophage, was determined by aligning the sequence of the intergenic region between *ejl* and *orf1*, as determined from the mature EJ-1 phage genome, to that of *attR* and the complete *S. pneumoniae* genome (Fig. 3). A 22-nt long core region could be defined. The assumption that the EJ-1 prophage was integrated in this region was confirmed by PCR amplification and sequencing of the DNA from strain 101 by using oligonucleotide primers designed to match *hlpA* (or the gene *degV* located immediately upstream in the pneumococcal genome) and *orf1*. However, attempts to identify the gene(s) located to the right of the prophage in the genome of the lysogenic strain 101 were unsuccessful. Nevertheless, it is noteworthy that almost immediately downstream of *attB*, a RUP_B2 element was found in the pneumococcal genome (Fig. 3). RUP elements appear to be involved in genome rearrangements (Oggioni and Claverys, 1999). Consequently, it is conceivable that the lack of similarity downstream of *attB* between the sequenced pneumococcal genomes and that of strain 101 may be due to a genomic rearrangement arising in the evolutionary history of this strain. Besides, and as shown in Fig. 3, several *S. pyogenes* prophages are also found downstream of *hlpA* although the surrounding genes are different between the GAS and *S. pneumoniae* genomes.

Identification of some structural proteins of EJ-1 by N-terminal aa sequencing

Purified EJ-1 virions were subjected to SDS-PAGE and up to 18 protein bands were observed although only the four major ones contained enough material to be employed for N-terminal aa sequence determination (Fig. 4A). This analysis revealed that these proteins were gp47 (most likely the major head protein, see Table 2), gp53, gp54, and gp65, and that all lack the first methionine, in agreement with previous rules (Ben-Bassat et al., 1987). To get further insight into the functions of gp53, gp54, and eventually other EJ-1 gene products, we also determined the N-terminal aa sequences of several proteins of the defective myovirus PBSX and compared them to those of PBSX protein bands whose function had been proposed on the basis of comparisons between the SDS-PAGE profile of complete phage particles and those from purified tails (Thurm and Garro, 1975a), tail sheaths

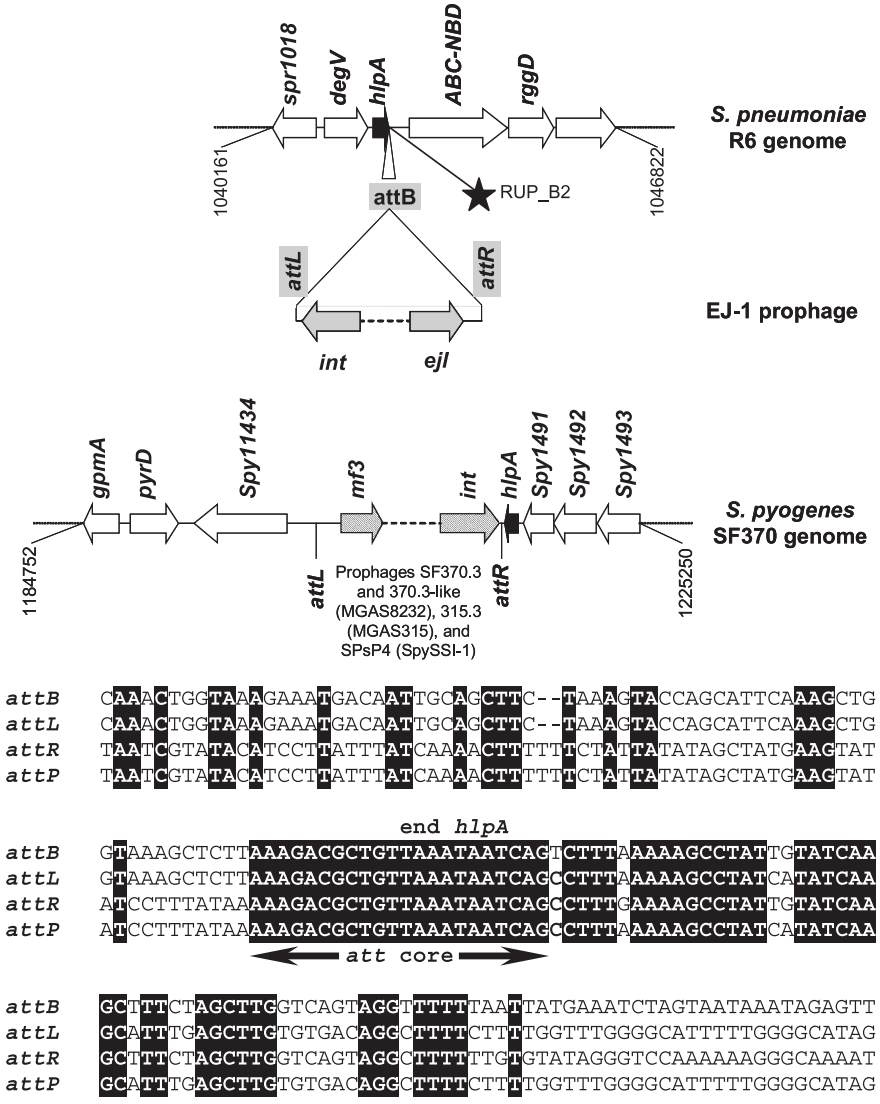

Fig. 3. Gene organization around the EJ-1 attachment site, and nucleotide sequence of *attB* (strain R6), *attP*, *attL*, and *attR*. The *hlpA* gene is depicted as a solid arrow, and the star indicates the location of the RUP element. Identical bases in all sequences are boxed in the multiple alignment shown at the bottom of the figure. The nucleotide sequences flanking the *attB* and *attR* sites correspond to the pneumococcal R6 genome (Hoskins et al., 2001) and to a partial EJ-1 sequence (Díaz et al., 1992a), respectively. The *attR*/*attL* sequence was also compared to that of the complete genome of the *S. pneumoniae* strain TIGR4 (Tettelin et al., 2001) as well as to the preliminary nt sequences of the pneumococcal strains 670 (http://www.tigrblast.tigr.org/ufmg) and Spain^{23F}-1 (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi?). An unfinished genomic sequence of the *S. mitis* type strain was also searched (http://www.tigrblast.tigr.org/ufmg). A near identical sequence located at the 3′ end of *hlpA* was found in every case (not shown).

preparations (Cremers et al., 1978; Steensma, 1981), or by using PBSX mutants affected in head or tail assembly (Thurm and Garro, 1975b), combined with electron microscopy. Six major PBSX bands could be assigned to the corresponding gene products (Fig. 4). Our results showed that the nine N-terminal aa of XkdG (34 614 Da) are removed proteolytically leaving a product of smaller molecular mass (33 445 Da) (XkdG*). Mauël and Karamata (1984) showed that the 36 kDa purported main capsid protein of PBSX was processed to a 35 kDa product. These and earlier data (see above) agree with the assumption that *xkdG* is the gene encoding the main capsid protein of PBSX. Likewise, the mature form of XkdF (XkdF*), which has only 65 aa residues, results from the elimination of the 210 N-terminal aa residues of XkdF and may correspond to the minor capsid protein (X7) previously reported (Thurm and Garro, 1975a, 1975b). Whether the same protease is responsible for the processing of XkdF and XkdG is not known, but it is interesting to note that the aa sequence flanking the cleavage site appears to be similar in both proteins (LKRLERIEK↓ in XkdF, and MRNQEIIRK↓ in XkdG). Whether the protease involved in the processing of those proteins is phage- or host-encoded is not known although a PBSX gene putatively coding for a protease could not be identified by sequence comparisons. The OpdA protease that cleaves the phage P22 gene 7 protein P22 is the only

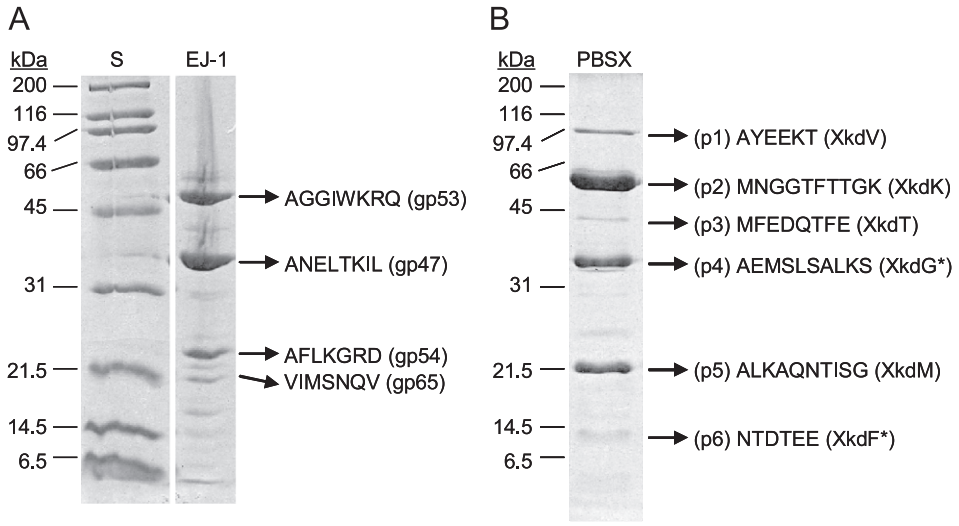

Fig. 4. SDS-PAGE and N-terminal aa sequence of the structural proteins of EJ-1 (A) and PBSX (B). The PBSX gene products were named according to Krogh et al. (1996). XkdF* and XkdG* denote the processed forms of XkdF and XkdG, respectively. The molecular mass of the standards (S) is indicated.

host protease that has been shown to cleave a phage virion protein (Conlin et al., 1992) whereas the protease responsible for processing the neck appendage protein precursor (p12) of ϕ29 and the processing site(s) have not been determined (Meijer et al., 2001). In several lambdoid phages either from Gram-positive (e.g., Sfi21) or Gram-negative (e.g., HK97) bacteria, proteolytic cleavage of the major structural proteins by a phage-encoded protease had been reported previously (Desiere et al., 1999). It has been shown that the phage HK97-encoded protease (gp4) cleaves 103 aa residues from the N terminus of each gp5 capsid protein subunit (the so-called "delta" region) (Duda et al., 1995), and it is assumed that the delta domain fulfills a scaffolding function (Hendrix and Duda, 1998). Besides, it was also observed that the cleavage was related with a predicted high potential for a coiled-coil structure of the delta domain (Conway et al., 1995). Bioinformatic analysis using the GOR4 program (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_gor4.html) (not shown) strongly suggested that the observed cleavage site of XkdF (between aa 210 and 211) is located immediately downstream of a predicted, long α helix, whereas in XkdG, the cleavage site is within a predicted coiled-coil region.

N-terminal cleavage processing does not take place in the putative tail proteins of PBSX (XkdK, XkdM, XkdT, and XkdV). On the basis of the migration of those proteins in SDS-polyacrylamide gels and previous results reported by others (see above), we propose that XkdK and XkdM may correspond, respectively, to the sheath and core tail proteins. XkdT and XkdV may be presumably a minor tail protein and a tail fiber subunit, respectively. In agreement with this assumption and the sequence similarities found between the tail components of PBSX and EJ-1 (Table 2) gp53, gp54, gp61, and gp64 would correspond to the sheath, core, a minor tail protein, and a fiber subunit of the EJ-1 tail, respectively.

EJ-1, a mosaic myovirus from a LGP bacterium

As already mentioned, several approaches to phage classification based on genome sequencing have been proposed (Lawrence et al., 2002; Proux et al., 2002; Rohwer and Edwards, 2002). For example, Proux et al. (2002) suggested a method based on the analysis of a single, but prominent, phage cluster, either the packaging-head cluster or the tail genes (excluding tail fiber genes). EJ-1 appears to belong to the *pac*-site siphoviruses group when the genes coding for the packaging and head proteins are taken into consideration, although it appears to belong to myoviruses by genomic analysis of the tail genes. We propose that EJ-1 represents the prototype of a previously undescribed group of phages from LGP bacteria. The genomic organization of the *Shigella flexneri* phage SfV, which also belongs to the *Myoviridae* morphology family, is analogous to that reported for that of EJ-1 as the head and tail proteins are proposed to originate from different morphology groups, that is, the λ-like group of siphoviruses and the Mu-like myoviruses, respectively (Allison et al., 2003). The defective *Salmonella enterica* phage ST64B represents an additional similar example, although some of the tail genes of this phage are inactivated by insertions of fragments of other genes related to virulence (Mmolawa et al., 2003). These two examples illustrate on bridges that provide an evolutionary link between phages, like Mu and λ, that do not have detectable sequence similarity. More genome sequences of myoviruses infecting LGP bacteria will be required before we can ascertain whether EJ-1 is unique or represents the first example of a large series of phages.

Phage gene accretion and virulence

The analysis of the region between the lysin and the integrase genes revealed an important difference to our

current knowledge on the organization of this region in temperate phages or prophages from LGP bacteria. With the notable exception of the *S. pneumoniae* phage MM1 recently described (Obregón et al., 2003a), all phages encode at least one gene in this region (Desiere et al., 2002), and in phages from pathogenic streptococci, many of them appear to be involved in virulence (Canchaya et al., 2003). However, a partial sequence of the genome of *Streptococcus canis* temperate phage ϕsc1 has revealed that it is mostly related to phage MM1 except that the latter does not carry *smc*, a gene coding for a mitogen responsible for canine streptococcal toxic shock syndrome and necrotizing fasciitis, that is inserted into a gene putatively encoding a minor tail protein and transcribed from a different reading frame. On the basis of differences in its G + C content, the *smc* gene must have been acquired by ϕsc1 from an unknown source, providing a noticeable example of phage gene accretion (Ingrey et al., 2003).

Although sequence analyses of the EJ-1 genome failed to reveal any gene potentially involved in host virulence, it should be mentioned that a recent report identified XkdG, XkdK, and XkdM as forming part of the extracellular proteome of *B. subtilis* (about 200 visible proteins), and proposed that these PBSX-related proteins were secreted through a holin-like pathway (Antelmann et al., 2001), most probably as a consequence of spontaneous phage induction (Shingaki et al., 2003). The extracellular presence and N-terminal processing of XkdG in *B. subtilis* cultures had been also previously noted (Hirose et al., 2000). Interestingly, an attempt to investigate toxin export kinetics and to identify other extracellular proteins and virulence factors of *Clostridium difficile* (an intestinal pathogen causing diarrhea and other complications) also identified a protein highly similar to XkdK (the homolog of gp53 from EJ-1) among the extracellular proteins secreted by *C. difficile* strain VPI 10463 when growing under conditions that allow high toxin production (Mukherjee et al., 2002). It is interesting to underline that only about 15 extracellular proteins were observed in this strain and that, apparently, neither XkdK nor the toxins showed altered levels after addition of mitomycin C. Whether gp53 and other PBSX-related proteins encoded by phage EJ-1 are also excreted to the medium by cultures of the lysogenic strain 101 and assist the virulence of the host deserves further research. In bacteria of noticeable clinical importance like the pneumococcus where an abundant presence of lytic and temperate phages has been proposed (Ramirez et al., 1999), a precise role of phage genes in bacterial infection and human disease remains to be established. Taking into account the fact that phages acquire genes both from other phages and their host, it is possible that their pervasively mosaic nature might contribute to bacterial pathogenicity. However, due to the great diversity of phage populations, the characterization of more genomes is required to fully address to these questions. The genomic exploration of the first myovirus (EJ-1) infecting LGP bacteria, a bacterial group that embraces important pathogens (as *S. pneumoniae* or *S. pyogenes*) as well as microorganisms of remarkable industrial importance (as *S. thermophilus*, lactococci, or lactobacilli), represents an important step in this direction.

Materials and methods

Bacterial strains, plasmids, phage purification, and growth conditions

We used strain 101, an atypical strain of *S. pneumoniae* (Díaz et al., 1992a, 1992b; Obregón et al., 2002), and the type strains of *S. mitis* (NCTC 12261) and *S. oralis* (NCTC 11427). *S. pneumoniae* 3870 was used as a source of the *ant* gene (see below) (Balsalobre et al., 2003). *Escherichia coli* DH5α (Hanahan, 1983) served as the host for recombinant plasmids harboring EJ-1 DNA fragments. *E. coli* was grown in Luria–Bertani (LB) medium (Sambrook et al., 1989) and *S. pneumoniae* in C medium (Lacks and Hotchkiss, 1960) supplemented with yeast extract (0.8 mg/ml; Difco Laboratories) (C + Y medium) or in Todd–Hewitt broth. The procedure for genetic transformation of *E. coli* has been previously described (Sambrook et al., 1989). Plasmid pJCP191 carries the *pnl* gene encoding the pneumococcal pneumolysin (Taira et al., 1989). Phage EJ-1 was prepared from strain 101 treated with mitomycin C (75 ng/ml) in the dark, and after lysis of the culture, the phage was purified by two cycles of CsCl density gradient centrifugation (Díaz et al., 1992a). Phage DNA was prepared by treatment of purified phage preparations with SDS and proteinase K (Díaz et al., 1992a). *B. subtilis* 168 was the strain used for the preparation of the defective, temperate phage PBSX following mitomycin C induction (75 ng/ml) (Mauël and Karamata, 1984). Plasmid pUC18 has been described (Yanisch-Perron et al., 1985).

PCR amplification, cloning, Southern blotting, and nt sequencing

Chromosomal DNA from streptococcal species was obtained as described previously (Fenoll et al., 1994). Routine DNA manipulations were performed essentially as described (Sambrook et al., 1989). DNA fragments were purified by using the Geneclean II kit (Bio 101). EJ-1 DNA was digested with appropriate restriction enzymes (*Acc*I, *Apo*I, *Eco*47III, *Eco*RV, *Hin*dIII, *Hin*P1, or *Xmn*I) and cloned into pUC18. The nt sequence was determined by the dideoxy chain-termination method (Sanger et al., 1977) with an automated Abi Prism 3700™ DNA sequencer (Applied Biosystems). Gaps in the sequence were filled using PCR-amplified segments or direct sequencing using the whole phage DNA as a template. The left end of the prophage was PCR-amplified using DNA from strain 101 and appropriate oligonucleotide primers designed to match the *hlpA* pneumococcal gene or *orf1* gene from EJ-1 (see above). The correctness of the sequence assembly was confirmed by

comparing the predicted map from this sequence with that experimentally obtained by using restriction enzymes on the phage DNA. Multilocus sequence typing (MLST) was carried out as described elsewhere (Enright and Spratt, 1998). We also used primers 63f/1387r to amplify and sequence the 16S rRNA gene (Marchesi et al., 1998), G-GaluP/F-GalUR for the determination of a partial nt sequence of the informative *galU* allele of strain 101 (Mollerach and García, 2000), and antUP/antDOWN to amplify the *ant* gene (Balsalobre et al., 2003). All primers for PCR amplification and nt sequencing were synthesized in-house on a Beckman model Oligo 1000 M synthesizer. DNA probes were labeled with the DIG Luminiscent Detection Kit (Boehringer Mannheim). Southern blots and hybridizations were carried out according to the manufacturer's instructions.

Data analysis

DNA and protein sequences were analyzed with the Genetics Computer Group (GCG) software package (version 10.0) (Devereux et al., 1984). PEDs (estimated number of substitutions per 100 bases) were determined using the DISTANCES program with the correction adequate to each case. Multiple sequence alignments were created with PILEUP. Putative promoters and transcription terminators were identified by using FINDPATTERNS and FOLDRNA, respectively. Moreover, the bioinformatic programs available at The Deambulum (http://www.infobiogen.fr/services/deambulum) and at the Institut Pasteur website (http://www.pasteur.fr/externe) were also employed. Sequence comparisons used the EMBL/SWALL databases and the FASTA (Pearson, 1990) or BLAST (Altschul et al., 1997) programs.

Miscellaneous techniques

CsCl-purified phage preparations were negatively stained with 1% uranyl acetate in carbon-reinforced, Formvar-coated copper grids (300 mesh) as previously described (Kay, 1976). Micrographs were taken on a LEO 910 transmission electron microscope working at 80 kV. SDS-PAGE was carried out with the buffer system described by Laemmli (1970) in 10% or 12.5% polyacrylamide gels, and protein bands were visualized by staining with Coomassie brilliant blue R250. N-terminal sequence analyses were carried out according to a published procedure (Speicher, 1994). Anti-EJ-1 serum was prepared by repeated injections (Lam and Mutharia, 1994) of purified phage preparations, as previously reported (López et al., 1977).

Nucleotide sequence accession numbers

The nt sequences determined in this study have been deposited in the EMBL/GeneBank/DDBJ databases. The 16S *rrna*, *aroE*, *ddl*, *galU*, *gdh*, *recP*, and *spi* alleles have been assigned the accession numbers AJ608938–AJ608944, and the complete nt sequence of the EJ-1 prophage is updated as accession number AJ609634.

Acknowledgments

We are grateful to P. García for helpful comments and for the critical reading of the manuscript, and to A.G. de la Campa for her advice on the amplification of the *ant* gene. We thank E. Cano and M. Carrasco for skilful technical assistance. This work was supported by grants from the Dirección General de Investigación Científica y Técnica (BCM2000-1002) and from Redes Temáticas de Investigación Cooperativa (G03/103 and C03/14) (Ministerio de Sanidad y Consumo). P. Romero is the recipient of a FPI fellowship from Ministerio de Ciencia y Tecnología.

References

Ackermann, H.W., 2003. Bacteriophage observations and evolution. Res. Microbiol. 154, 245–251.

Allison, G., Angeles, D.C., Huan, P.T., Verma, N.K., 2003. Morphology of temperate bacteriophage SfV and characterization of the DNA packaging and capsid genes: the structural genes evolved from two different phage families. Virology 308, 114–127.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Antelmann, H., Tjalsma, H., Voigt, B., Ohlmeier, S., Bron, S., van Dijl, J.M., Hecker, M., 2001. A proteomic view on genome-based signal peptide predictions. Genome Res. 11, 1484–1502.

Bacot, C.M., Reeves, R.H., 1991. Novel tRNA gene organization in the 16S–23S intergenic spacer of the *Streptococcus pneumoniae* rRNA gene cluster. J. Bacteriol. 173, 4234–4236.

Balsalobre, L., Ferrándiz, M.J., Liñares, J., Tubau, F., de la Campa, A.G., 2003. Viridans group streptococci are donors in horizontal transfer of topoisomerase IV genes to *Streptococcus pneumoniae*. Antimicrob. Agents Chemother. 47, 2072–2081.

Ben-Bassat, A., Bauer, K., Chang, S.-Y., Myambo, K., Boosman, A., Chang, S., 1987. Processing of the initiation methionine from proteins: properties of the *Escherichia coli* methionine aminopeptidase and its gene structure. J. Bacteriol. 169, 751–757.

Bernheimer, H.P., 1977. Lysogeny in pneumococci freshly isolated from man. Science 195, 66–68.

Brüssow, H., 2001. Phages of dairy bacteria. Annu. Rev. Microbiol. 55, 283–303.

Brüssow, H., Desiere, F., 2001. Comparative phage genomics and the evolution of *Siphoviridae*: insights from dairy phages. Mol. Microbiol. 39, 213–222.

Brüssow, H., Hendrix, R.W., 2002. Phage genomics: small is beautiful. Cell 108, 13–16.

Bruttin, A., Desiere, F., Lucchini, S., Foley, S., Brüsow, H., 1997. Characterization of the lysogeny module from the temperate *Streptococcus thermophilus* bacteriophage ϕSfi11. Virology 233, 136–148.

Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M.L., Brüssow, H., 2003. Phage as agents of lateral gene transfer. Curr. Opin. Microbiol. 6, 417–424.

Conlin, C.A., Vimr, E.R., Miller, C.G., 1992. Oligopeptidase a is required for normal phage P22 development. J. Bacteriol. 174, 5869–5880.

Conway, J.F., Duda, R.L., Cheng, N., Hendrix, R.W., Steven, A.C., 1995.

Proteolytic and conformational control of virus capsid maturation: the bacteriophage HK97 system. J. Mol. Biol. 253, 86–99.
Cremers, A.F.M., de Steensma, H.Y., Mellema, J.E., 1978. The quaternary structure of the sheaths of defective phages similar to PBS X. Eur. J. Biochem. 89, 389–395.
Desiere, F., Lucchini, S., Brüssow, H., 1999. Comparative sequence analysis of the DNA packaging, head, and tail morphogenesis modules in the temperate cos-site *Streptococcus thermophilus* bacteriophage Sfi21. Virology 260, 244–253.
Desiere, F., Pridmore, R.D., Brüssow, H., 2000. Comparative genomics of the late gene cluster from *Lactobacillus* phages. Virology 275, 294–305.
Desiere, F., Mahanivong, C., Hillier, A.J., Chandry, P.S., Davidson, B.E., Brüssow, H., 2001a. Comparative genomics of lactococcal phages: insight from the complete genome sequence of *Lactococcus lactis* phage BK5-T. Virology 283, 240–252.
Desiere, F., McShan, W.M., van Sinderen, D., Ferretti, J.J., Brüssow, H., 2001b. Comparative genomics reveals close genetic relationships between phages from dairy bacteria and pathogenic streptococci: evolutionary implications for prophage–host interactions. Virology 288, 325–341.
Desiere, F., Lucchini, S., Canchaya, C., Ventura, M., Brüssow, H., 2002. Comparative genomics of phages and prophages in lactic acid bacteria. Antonie van Leeuwenhoek 82, 73–91.
Devereux, J., Haeberli, P., Smithies, O., 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12, 387–395.
Díaz, E., López, R., García, J.L., 1992a. EJ-1, a temperate bacteriophage of Streptococcus pneumoniae with a *Myoviridae* morphotype. J. Bacteriol. 174, 5516–5525.
Díaz, E., López, R., García, J.L., 1992b. Role of the major pneumococcal autolysin in the atypical response of a clinical isolate of *Streptococcus pneumoniae*. J. Bacteriol. 174, 5508–5515.
Díaz, E., Munthali, M., Lunsdorf, H., Höltje, J.-V., Timmis, K.N., 1996. The two-step lysis system of pneumococcal bacteriophage EJ-1 is functional in gram-negative bacteria: triggering of the major pneumococcal autolysin in *Escherichia coli*. Mol. Microbiol. 19, 667–681.
Duda, R.L., Hempel, J., Michel, H., Shabanowitz, J., Hunt, D., Hendrix, R.W., 1995. Structural transitions during bacteriophage HK97 head assembly. J. Mol. Biol. 247, 618–635.
Enright, M.C., Spratt, B.G., 1998. A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. Microbiology 144, 3049–3060.
Fenoll, A., Muñoz, R., García, E., de la Campa, A.G., 1994. Molecular basis of the optochin-sensitive phenotype of pneumococcus: characterization of the genes encoding the F0 complex of the *Streptococcus pneumoniae* and *Streptococcus oralis* H^+-ATPases. Mol. Microbiol. 12, 587–598.
Hanahan, D., 1983. Studies on transformation of *E. coli* with plasmids. J. Mol. Biol. 166, 557–580.
Haro, A., Vélez, M., Goormaghtigh, E., Lago, S., Vázquez, J., Andreu, D., Gasset, M., 2003. Reconstitution of holin activity with a synthetic peptide containing the 1–32 sequence region of Ejh, the EJ-1 phage holin. J. Biol. Chem. 278, 3936–3939.
Hendrix, R.W., 2003. Bacteriophage genomics. Curr. Opin. Microbiol. 6, 506–511.
Hendrix, R.W., Duda, R.L., 1998. Bacteriophage HK97 head assembly: a protein ballet. Adv. Virus Res. 50, 235–288.
Hirose, I., Sano, K., Shioda, I., Kumano, M., Nakamura, K., Yamane, K., 2000. Proteome analysis of *Bacillus subtilis* extracellular proteins: a two-dimensional protein electrophoretic study. Microbiology 146, 65–75.
Hoskins, J., Alborn, W.E., Arnold, J., Blaszczak, L.C., Burgett, S., DeHoff, B.S., Estrem, S.T., Fritz, L., Fu, D.-J., Fuller, W., Geringer, C., Gilmour, R., Glass, J.S., Khoje, H., Kraft, A.R., Lagace, R.E., LeBlanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P., McAhren, S.M., McHenney, M., McLeaster, K., Mundy, C.W., Nicas, T.I., Norris, F.H., O'gara, M., Peery, R.B., Robertson, G.T., Rockey, P., Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C.A., Baltz, R.H., Jaskunas, R., Rosteck, P.R.J., Skatrud, P.L., Glass, J.I., 2001. Genome of the bacterium *Streptococcus pneumoniae* strain R6. J. Bacteriol. 183, 5709–5717.
Ingrey, K.T., Ren, J., Presscott, J.F., 2003. A fluoroquinolone induces a novel mitogen-encoding bacteriophage in *Streptococcus suis*. Infect. Immun. 71, 3028–3033.
Jado, I., López, R., García, E., Fenoll, A., Casal, J., García, P., 2003. Phage lytic enzymes as therapy of antibiotic-resistant *Streptococcus pneumoniae* infection in a murine sepsis model. J. Antimicrob. Chemother. 52, 967–973.
Juhala, R.J., Ford, M.E., Duda, R.L., Youlton, A., Hatfull, G.F., Hendrix, R.W., 2000. Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. J. Mol. Biol. 299, 27–51.
Jukes, T.H., Cantor, C.R., 1969. Evolution of protein molecules. In: Munro, H.N. (Ed.), Mammalian Protein Metabolism, vol. II. Academic Press Inc, NY, pp. 21–132.
Kawamura, Y., Whiley, R.A., Shu, S.-E., Ezaki, T., Hardie, J.M., 1999. Genetic approaches to the identification of the mitis group within the genus *Streptococcus*. Microbiology 145, 2605–2613.
Kay, D., 1976. Electron microscopy of small particles, macromolecular structures and nucleic acids. Methods Microbiol. 9, 177–215.
Krogh, S., O'Reilly, M., Nolan, N., Devine, K.M., 1996. The phage-like element PBSX and part of the skin element, which are resident at different locations on the *Bacillus subtilis* chromosome, are highly homologous. Microbiology 142, 2031–2040.
Lacks, S., Hotchkiss, R.D., 1960. A study of the genetic material determining an enzyme activity in *Pneumococcus*. Biochim. Biophys. Acta 39, 508–517.
Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227, 680–685.
Lam, J.S., Mutharia, L.M., 1994. Antigen–antibody reactions. In: Gerhardt, P., Murray, R.G.E., Wood, W.A., Krieg, N.R. (Eds.), Methods for General and Molecular Bacteriology. American Society for Microbiology, Washington, DC.
Lawrence, J.G., Hatfull, G.F., Hendrix, R.W., 2002. Imbroglios of viral taxonomy: genetic exchange and failings of phenetic approaches. J. Bacteriol. 184, 4891–4905.
Loeffler, J.M., Nelson, D., Fischetti, V.A., 2001. Rapid killing of *Streptococcus pneumoniae* with a bacteriophage cell wall hydrolase. Science 294, 2170–2172.
Loessner, M.J., Inman, R.B., Lauer, P., Calendar, R., 2000. Complete nucleotide sequence, molecular analysis and genome structure of bacteriophage A118 of *Listeria monocytogenes*: implications for phage evolution. Mol. Microbiol. 35, 324–340.
López, R., Ronda, C., Tomasz, A., Portolés, A., 1977. Properties of "diplophage": a lipid-containing bacteriophage. J. Virol. 24, 201–210.
Lucchini, S., Desiere, F., Brüssow, H., 1999. Comparative genomics of *Streptococcus thermophilus* phage species supports a modular evolution theory. J. Virol. 73, 8647–8656.
Marchesi, J.R., Sato, T., Weightman, A.J., Martin, T.A., Fry, J.C., Hiom, S.J., Wade, W.G., 1998. Design and evaluation of useful bacterium-specific PCR primers that amplify genes coding for bacterial 16S rRNA. Appl. Environ. Microbiol. 64, 795–799.
Martín, A.C., López, R., García, P., 1996. Analysis of the complete nucleotide sequence and functional organization of the genome of *Streptococcus pneumoniae* bacteriophage Cp-1. J. Virol. 70, 3678–3687.
Mauël, C., Karamata, D., 1984. Characterization of proteins induced by mitomycin C treatment of *Bacillus subtilis*. J. Virol. 49, 806–812.
Meijer, W.J.J., Horcajadas, J.A., Salas, M., 2001. ϕ29 family of phages. Microbiol. Mol. Biol. Rev. 65, 261–287.
Mmolawa, P.T., Schmieger, H., Heuzenroeder, M.W., 2003. Bacteriophage ST64B, a genetic mosaic of genes from diverse sources isolated from *Salmonella enterica* serovar Typhimurium DT 64. J. Bacteriol. 185, 6481–6485.
Mollerach, M., García, E., 2000. The *galU* gene of *Streptococcus pneumoniae* that codes for a UDP-glucose pyrophosphorylase is highly poly-

morphic and suitable for molecular typing and phylogenetic studies. Gene 260, 77–86.

Mukherjee, K., Karlsson, S., Burman, L.G., Akerlund, T., 2002. Proteins released during high toxin production in *Clostridium difficile*. Microbiology 148, 2245–2253.

Nelson, D., Schuch, R., Zhu, S., Tscherne, D.M., Fischetti, V.A., 2003. Genomic sequence of C_1, the first streptococcal phage. J. Bacteriol. 185, 3325–3332.

Obregón, V., García, P., García, E., Fenoll, A., López, R., García, J.L., 2002. Molecular peculiarities of the *lytA* gene isolated from clinical pneumococcal strains that are bile-insoluble. J. Clin. Microbiol. 40, 2545–2554.

Obregón, V., García, J.L., García, E., López, R., García, P., 2003a. Genome organization and molecular analysis of the temperate bacteriophage MM1 of *Streptococcus pneumoniae*. J. Bacteriol. 185, 2362–2368.

Obregón, V., García, P., López, R., García, J.L., 2003b. VO1, a temperate bacteriophage of the type 19A multiresistant epidemic 8249 strain of *Streptococcus pneumoniae*: analysis of variability of lytic and putative C5 methyltransferase genes. Microb. Drug Resist. 9, 7–15.

Oggioni, M.R., Claverys, J.P., 1999. Repeated extragenic sequences in procaryotic genomes: a proposal for the origin and dynamics of the RUP element in *Streptococcus pneumoniae*. Microbiology 145, 2647–2653.

Pearson, W.R., 1990. Rapid and sensitive sequence comparison with FASTP and FASTA. Methods Enzymol. 183, 63–98.

Plunkett III, G., Rose, D.J., Durfee, T.J., Blattner, F.R., 1999. Sequence of Shiga toxin 2 phage 933W from *Escherichia coli* O157:H7: shiga toxin as a phage late-gene product. J. Bacteriol. 181, 1767–1778.

Proux, C., van Sinderen, D., Suarez, J., Garcia, P., Ladero, V., Fitzgerald, G.F., Desiere, F., Brüssow, H., 2002. The dilemma of phage taxonomy illustrated by comparative genomics of Sfi21-like *Siphoviridae* in lactic acid bacteria. J. Bacteriol. 184, 6026–6036.

Ramirez, M., Severina, E., Tomasz, A., 1999. A high incidence of prophage carriage among natural isolates of *Streptococcus pneumoniae*. J. Bacteriol. 181, 3618–3625.

Rohwer, F., Edwards, R., 2002. The phage proteomic tree: a genome-based taxonomy for phage. J. Bacteriol. 184, 4529–4535.

Sáiz, J.L., López-Zumel, C., Monterroso, B., Varea, J., Arrondo, J.L.R., Iloro, I., García, J.L., Laynez, J., Menéndez, M., 2002. Characterization of Ejl, the cell-wall amidase coded by the pneumococcal bacteriophage Ej-1. Protein Sci. 11, 1788–1799.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. 74, 5463–5467.

Severina, E., Ramirez, M., Tomasz, A., 1999. Prophage carriage as a molecular epidemiological marker in *Streptococcus pneumoniae*. J. Clin. Microbiol. 37, 3308–3315.

Shingaki, R., Kasahara, Y., Inoue, T., Kokeguchi, S., Fukui, K., 2003. Chromosome DNA fragmentation and excretion caused by defective prophage gene expression in the early-exponential-phase culture of *Bacillus subtilis*. Can. J. Microbiol. 49, 313–325.

Speicher, D.W., 1994. Methods and strategies for the sequence analysis of proteins on PVDF membranes. Methods 6, 262–273.

Stackebrandt, E., Frederiksen, W., Garrity, G.M., Grimont, P.A.D., Kämpfer, P., Maiden, M.C.J., Nesme, X., Roselló-Mora, R., Swings, J., Trüper, H.G., Vauterin, L., Ward, A.C., Whitman, W.B., 2002. Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. Int. J. Syst. Evol. Microbiol. 52, 1043–1047.

Stanley, E., Walsh, L., van der Zwet, A., Fitzgerald, G.F., van Sinderen, D., 2000. Identification of four loci isolated from two *Streptococcus thermophilus* phage genomes responsible for mediating bacteriophage resistance. FEMS Microbiol. Lett. 182, 271–277.

Steensma, H.Y., 1981. Effect of defective phages on the cell membrane of *Bacillus subtilis* and partial characterization of a phage protein involved in killing. J. Gen. Virol. 56, 275–286.

Taira, S., Jalonen, E., Paton, J.C., Sarvas, M., Runeberg-Nyman, K., 1989. Production of pneumolysin, a pneumococcal toxin, in *Bacillus subtilis*. Gene 77, 211–218.

Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D., Peterson, S., Heidelber, J., DeBoy, R.T., Haft, D.H., Dodson, R.J., Durkin, A.S., Gwinn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D., Umayam, L.A., White, O., Salzberg, S.L., Lewis, M.R., Radune, D., Holtzapple, E., Khouri, H., Wolf, A.M., Utterback, T.R., Hansen, C.L., McDonald, L.A., Feldblyum, T.V., Angiuoli, S., Dickinson, T., Hickey, E.K., Holt, I.E., Loftus, B.J., Yang, F., Smith, H.O., Venter, J.C., Dougherty, B.A., Morrison, D.A., Hollingshead, S.K., Fraser, C.M., 2001. Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. Science 293, 498–506.

Thurm, P., Garro, A.J., 1975a. Bacteriophage-specific protein synthesis during induction of the defective *Bacillus subtilis* bacteriophage PBSX. J. Virol. 16, 179–183.

Thurm, P., Garro, A.J., 1975b. Isolation and characterization of prophage mutants of the defective *Bacillus subtilis* bacteriophage PBSX. J. Virol. 175, 184–191.

Vybiral, D., Takác, M., Loessner, M., Witte, A., von Ahsen, U., Bläsi, U., 2003. Complete nucleotide sequence and molecular characterization of two lytic *Staphylococcus aureus* phages: 44AHJD and P68. FEMS Microbiol. Lett. 219, 275–283.

Yanisch-Perron, C., Vieira, J., Messing, J., 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33, 103–119.